Innovation in Residency Selection: The AAMC Standardized Video Interview

Steven B. Bird, MD, H. Gene Hern, MS, MD, Andra Blomkalns, MD, Nicole M. Deiorio, MD, Yolanda Haywood, MD, Katherine M. Hiller, MD, MPH, Dana Dunleavy, PhD, and Keith Dowd, MA

Abstract

Purpose

Innovative tools are needed to shift residency selection toward a more holistic process that balances academic achievement with other competencies important for success in residency. The authors evaluated the feasibility of the AAMC Standardized Video Interview (SVI) and evidence of the validity of SVI total scores.

Method

The SVI, developed by the Association of American Medical Colleges, consists of six questions designed to assess applicants' interpersonal and communication skills and knowledge of professionalism. Study 1 was conducted in 2016 for research purposes. Study 2 was an operational pilot administration in 2017; SVI data were available for use in residency selection by emergency medicine programs for the 2018 application cycle. Descriptive statistics, correlations, and standardized mean differences were used to examine data.

Results

Study 1 included 855 applicants; Study 2 included 3,532 applicants. SVI total scores were relatively normally distributed. There were small correlations between SVI total scores and United States Medical Licensing Examination Step exam scores, Alpha Omega Alpha Honor Medical Society membership, and Gold Humanism Honor Society membership. There were no-to-small group differences in SVI total scores by gender and race/ethnicity, and small-to-medium differences by applicant type.

Conclusions

Findings provide initial evidence of the validity of SVI total scores and suggest that these scores provide different information than academic metrics. Use of the SVI, as part of a holistic screening process, may help program directors widen the pool of applicants invited to in-person interviews and may signal that programs value interpersonal and communication skills and professionalism.

Residency selection is a critical undertaking that allows residency programs to identify applicants who are a good fit for a particular medical specialty and/or program. On the whole, the current selection process works well in identifying applicants who are academically prepared to begin training and who will be successful in passing board exams.[1] However, it could benefit from incorporating new tools to assess applicants' professionalism and/or interpersonal and communication skills because those skills are critical for patient care.[2–7] It could also benefit from implementing new tools to help program directors balance United States Medical Licensing Examination (USMLE) Step exam scores in the selection process and signal that programs value broad preparation and diversity.[8,9]

In a 2016 survey conducted by the Association of American Medical Colleges (AAMC),[10] program directors reported being least satisfied with information currently available in the pre-interview screening stage about applicants' interpersonal and communication skills and professionalism. In addition, 60% of respondents reported that one of their top three challenges in the selection process is a lack of reliable information about applicants' personal competencies. Although tools that assess these competencies (e.g., Medical School Performance Evaluation [MSPE], personal statements, and letters of recommendation) are available, most are difficult to use in a high-volume context because of low reliability, the inability to compare candidates across schools, and the time-intensive nature of reading these documents.[7] Data from a 2017 AAMC survey of program directors from multiple specialties showed that program staff spent an average of 196 person-hours reviewing applications each year (or the equivalent of one full-time employee working 40 hours per week for 4.9 weeks).[11] As a result, many program directors indicated that they defer detailed assessment of personal competencies until the in-person interview—after the majority of applicants have been excluded largely on the basis of their academic achievements.

To address calls for innovative residency selection tools that can help mitigate overreliance on USMLE scores, identify applicants with strengths in important nonacademic competencies, and signal that the medical education community values breadth of clinical skills and diversity in the U.S. physician workforce,[2–9,12] some specialties have begun to develop alternative assessments. These include the electronic Standardized Letter of Evaluation (eSLOE) in emergency medicine (EM),[13,14] video interviews in orthopedic surgery and urology,[15] and telephone interviews in otolaryngology.

For such tools to be used and effective, they must be easy to use in a high-volume context and available for use at the first step in the application screening process. After reviewing alternative selection

Please see the end of this article for information about the authors.

The authors have informed the journal that they agree that both Steven B. Bird and H. Gene Hern completed the intellectual and other work typical of the work of the first author.

Correspondence should be addressed to Dana Dunleavy, Association of American Medical Colleges, 655 K St., Washington, DC 20001; telephone: (202) 862-6011; email: ddunleavy@aamc.org.

Acad Med. 2019;94:1489–1497.
First published online March 12, 2019
doi: 10.1097/ACM.0000000000002705
Copyright © 2019 by the Association of American Medical Colleges

tools used in employment[16] and higher education,[17] the AAMC determined that a structured video interview had the most potential to be useful in the residency selection context.

Research suggests that structured interviews have adequate validity in predicting job performance, result in smaller group differences than standardized tests, and are likely more resistant to coaching and deceit than tools such as personality inventories and biodata questionnaires.[18] Unfortunately, face-to-face structured interviews typically cannot be used for high-volume screening because of the expense. As such, video interviews may reduce costs while widening reach for large applicant pools. Although there is limited research on video interviews, initial findings show that compared with face-to-face interviews, on average, raters have higher levels of agreement evaluating video interviews,[19] applicants receive lower scores on technology-mediated videos,[19] and applicants have less positive reactions to technology-mediated interviews.[20] More research is needed to understand whether the advantages of increased structure that are well established in the face-to-face interview literature extend to video interviews.

We designed two studies to investigate the feasibility of a novel tool—the AAMC Standardized Video Interview (SVI)—for use in residency selection. Study 1 was conducted in 2016 for research purposes only, and Study 2 was conducted in 2017 during an operational pilot administration. In both studies, we evaluated the following: rater agreement; score distributions; performance differences by race/ethnicity, gender, and applicant type; and correlations with Electronic Residency Application Service (ERAS) data. Together, these studies allowed us to evaluate evidence of the validity of SVI total scores using the framework provided by the Standards for Educational and Psychological Testing jointly developed by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.[21]

Method

Each study was reviewed by the AAMC Human Subjects Research Protection Program and approved by the institutional review board of the American Institutes for Research (FWA00001666). Applicants provided consent for their personal data to be used in research when they completed the ERAS application and when they agreed to complete the SVI.

Study participants

Study 1 (2016 cohort). Applicants who indicated interest in applying to EM on the ERAS application or who applied to pediatrics, internal medicine, or general surgery for the ERAS 2017 cycle were invited to participate via email. Although we originally planned to include EM applicants only, we ultimately included more programs' applicants to increase the possible number of participants, to ensure that we had sufficient data for all analyses. The study was open from June 2016 to December 2016, and applicants received a $50 gift card code to Amazon.com for participating. We compared the demographic composition of the EM sample versus the EM applicant pool and found that the sample was representative with respect to race/ethnicity and gender (data not shown). There was a slightly higher proportion of attendees of U.S. MD-granting medical schools (US-MDs) in the EM sample than in the EM applicant population.

Applicants participating in the SVI were randomly assigned to one of four forms of interview questions. Each form consisted of a unique set of questions. Only results from forms 1 and 2 are reported because analyses of forms 3 and 4 had not been conducted at the time of writing. The overall SVI participation rate was 11.3% (1,760 completed/15,529 invited). Of those, 855 applicants (49%) were included in the final sample because they completed form 1 or 2.

Study 2 (2017 cohort). Applicants who indicated interest in applying to EM for the ERAS 2018 cycle were asked to complete the SVI as part of their application to EM residency programs. Applicants participating in the SVI were randomly assigned to one of multiple forms, and results from all forms are reported in this article. (The number of forms used cannot be shared because of concerns about the security and the integrity of SVI total scores.) The SVI was open from June 6 to July 31, 2017. Applicants were not required to participate in the SVI, but they were encouraged to do so, and administration was free of charge. Applicants were told that some programs planned to use SVI total scores in their selection process. The overall SVI participation rate was 84% (3,532 completed/4,229 invited). However, not all 3,532 applicants who completed the SVI and were included in this analysis applied to EM programs; the final result was that 85% of all EM applicants completed the SVI (3,469/4,060).

Table 1 provides a summary of both samples' characteristics.

SVI processes

The same processes were used to create the SVI questions and forms and to evaluate applicant responses in Study 1 and Study 2.

SVI. The SVI is an online, asynchronous interview designed to assess applicants' proficiency in two of the Accreditation Council for Graduate Medical Education (ACGME) competencies: interpersonal and communication skills and professionalism.[3] For the SVI, we renamed professionalism as "knowledge of professional behavior" to acknowledge that the SVI is not a direct observation of behavior but, rather, allows an inference of proficiency based on an applicant's description of past experiences or what he or she should do in hypothetical situations.

A team of experts in high-stakes assessments developed interview questions. Then, 20 subject matter experts from EM, internal medicine, pediatrics, neurology, obstetrics–gynecology, and general surgery reviewed each question for (1) relevance to the target competency and (2) potential for bias. Only questions that survived the expert review were retained. A mix of past-behavior and hypothetical questions was used to ensure that applicants who may not have had an opportunity to demonstrate some behaviors could respond.

After receiving an emailed invitation, applicants logged into the SVI site, where they completed technology checks, were given the opportunity to watch a brief introductory video, and could complete an unlimited number of practice questions. The practice questions were provided by the interview vendor and did not map to the target competencies. The

Table 1

Comparison of the AAMC Standardized Video Interview (SVI) Study 1 and Study 2 Samples[a]

Characteristic	Study 1 (2016 cohort), no. (% of 855)[b]	Study 2 (2017 cohort), no. (% of 3,532)[b]
Applied to at least one ACGME program		
Emergency medicine	481 (56)	3,469 (98)
Pediatrics	111 (13)	—
Obstetrics–gynecology	54 (6)	—
General surgery	104 (12)	—
Other specialty	118 (14)	—
Did not apply	24 (3)	63 (2)
Race/ethnicity[c]		
White	397 (47)	2,001 (57)
Black	90 (11)	247 (7)
Latino	63 (7)	286 (8)
Asian	218 (26)	613 (17)
Other	43 (5)	157 (4)
Did not report	59 (7)	283 (8)
Gender		
Male	516 (61)	2,311 (65)
Female	337 (39)	1,219 (35)
Did not report	2 (< 1)	2 (< 1)
Applicant type		
US-MD	441 (52)	2,062 (58)
DO	114 (13)	915 (26)
US-IMG	137 (16)	320 (9)
FMG	161 (19)	220 (6)
Unknown	2 (< 1)	15 (< 1)

Abbreviations: AAMC indicates Association of American Medical Colleges; ACGME, Accreditation Council for Graduate Medical Education; US-MD, attendee of a U.S. MD-granting medical school; DO, attendee of a DO-granting medical school; US-IMG, U.S. citizen attendee of an international medical school; FMG, non-U.S. citizen attendee of an international medical school; ERAS, Electronic Residency Application Service.

[a] In Study 1, conducted for research purposes, applicants participated in the SVI during 2016 for the ERAS 2017 cycle. These applicants indicated interest in applying to emergency medicine or applied to the other specialties indicated in this table. In Study 2, the operational pilot, applicants participated in the SVI in 2017 for the ERAS 2018 cycle. The pilot was limited to applicants who indicated interest in applying to emergency medicine. Participation in the SVI was optional and free of charge in both studies.

[b] Percentages may not sum to 100% because of rounding.

[c] For race/ethnicity, percentages do not sum to 100% because individuals who self-identified as white alone were classified as white, individuals who self-identified as black alone or in combination with other races (including white) were classified as black, and individuals who self-identified as Latino alone or in combination with other races (including white) were classified as Latino.

purpose of the practice questions was for applicants to get comfortable with using the SVI interface and completing a technology-mediated interview. Then, when applicants began the SVI, they were presented with a series of six questions (three targeting each competency) in text format. Applicants had up to 30 seconds to read each question and prepare a response. Once applicants were ready to respond (or the 30 seconds were up), they responded to the question verbally. Responses were recorded by the computer webcam. Applicants had up to three minutes to respond to each question. A sample question for each competency is provided below:

- Describe a situation in which you were successful in communicating a difficult message. How did you communicate the message? What was the outcome? (*interpersonal and communication skills*)
- Describe a situation in which you noticed a mistake or an error that had been made. What was the situation? What action did you take? What was the outcome? (*knowledge of professional behavior*)

Evaluation of responses. Applicants' responses to each question were evaluated using a scoring rubric designed for the appropriate target competency. Each scoring rubric ranged from a low of 1 (rudimentary) to a high of 5 (exemplary). For each proficiency level, raters were provided a general description and behavioral examples specific to each competency. (Specific behavioral examples included on the rubrics cannot be shared because of concerns about the security and the integrity of SVI total scores.) The rating scales were developed after a review of the ACGME competencies[3] and milestones for EM, pediatrics, internal medicine, surgery, and psychiatry,[22] with the help of EM program directors and faculty. Subject matter experts reviewed each example behavior for (1) relevance to the target competency, (2) potential for bias, and (3) placement at the appropriate proficiency level. These expert reviews provided validity evidence based on interview content.

A demographically diverse cohort of human resources professionals was contracted by the AAMC to rate applicants' responses. Rater training covered the EM trainee job, the two competencies, a standardized rating process, and unconscious bias. Program directors and faculty from academic EM programs watched, rated, and discussed actors' portrayals of several applicant responses to SVI questions and came to consensus on the rating for these responses. These video portrayals and consensus ratings were then used to train raters to identify what program directors were looking for in a response and how raters should interpret responses to meet the program director standard. Raters also participated in calibration activities in which they practiced making ratings and received feedback about their ratings. Although the content of the rater training was similar across studies, the length of training was increased from about 4 hours in Study 1 to 16 hours in Study 2. The training in Study 2 had an increased focus on using a structured process for evaluating responses and devoted more time to unconscious bias. There was no overlap in the raters used for Study 1 and Study 2. The order in which raters evaluated participants' responses was randomized to minimize effects of potential rater biases (e.g., order effects) on participants' SVI total scores.

Six raters were assigned to each applicant (i.e., one rater per question) to limit the influence any one rater had on an applicant's SVI total score. Ratings for each question were summed to create an SVI total score that ranged from 6 to 30. As with in-person interviews, raters made inferences about an applicant's proficiency level on the target competency and assigned a rating, using the rating scale, based on the applicant's description of his or her past behavior or of what he or she should do in a hypothetical situation.

Rater agreement was evaluated with calibration activities before the official rating period using an intraclass correlation ICC (2, k).[23] In Study 1, raters evaluated 120 complete interviews (2,160 total ratings), and rater agreement was ICC (2, k) = 0.81. In Study 2, raters evaluated 60 complete interviews (360 total ratings), and rater agreement was ICC (2, k) = 0.78. Data from these calibration activities provide validity evidence based on response processes.

Selection variables

Demographic information included applicants' self-reported gender, race/ethnicity, age, and applicant type (i.e., US-MD; U.S. citizen attendee of an international medical school [US-IMG]; non-U.S. citizen attendee of an international medical school [FMG]; and attendee of a DO-granting medical school [DO]). We included these data to examine potential differences in SVI total scores by demographic group. We expected small-to-medium differences.

Scores for applicants' first attempts on the USMLE Step 1, Step 2 Clinical Knowledge (CK), and Step 2 Clinical Skills (CS) exams were extracted from the ERAS application. We included these data to evaluate validity evidence based on relations with other variables. We expected no correlation between SVI total scores and Step 1 scores, and small correlations between SVI total scores and Step 2 CK and CS scores.

Data on membership in the Alpha Omega Alpha (AOA) Honor Medical Society, if available, were included in the analysis. Although medical schools have different processes, nomination and induction are largely based on students' academic accomplishments in medical school. We included these data, which were self-reported and extracted from the ERAS application, to evaluate validity evidence based on relations with other variables. We expected no correlation between SVI total scores and AOA membership.

Data on membership in the Gold Humanism Honor Society (GHHS), which recognizes students who demonstrate compassionate patient care and community service, were included in the analysis when available. Medical schools use different processes to select GHHS members. GHHS recommends using McCormack and colleagues'[24] peer-nomination survey to identify students at the end of the third year and select approximately 10% to 15% of each medical school class. We included these data, which were self-reported and extracted from the ERAS application, to evaluate validity evidence based on relations with other variables. We expected a small correlation between SVI total scores and GHHS membership.

Statistical analyses

The unit of analysis was the individual applicant. Data were linked using the applicant's AAMC ID. All analyses were conducted using SPSS version 19 (IBM Corp., Armonk, New York). Rater reliability across multiple raters was computed using an intraclass correlation ICC (2, k).[23] Descriptive statistics, including mean, standard deviation (SD), and total score distributions, were computed. SVI total scores were compared for participants from different demographic groups using descriptive statistics, evaluating the size of the difference with *t* tests and standardized mean differences (Cohen's *d*).[25] The relationships between SVI total scores and selection data were evaluated with Pearson correlations or point-biserial correlations.

Results

As shown in Figure 1, for both study cohorts, SVI total scores had similar means and SDs. The scores were approximately normally distributed, and raters used the full range of the rating scale.

Table 2 summarizes SVI total scores and standardized mean differences[25] by demographic group. The 2016 cohort (Study 1) included 855 participants, and SVI total scores ranged from 9 to 29 (mean [SD] = 18.7 [2.8]). The 2017 cohort (Study 2) included 3,532 participants, and SVI total scores ranged from 6 to 29 (mean [SD] = 19.1 [3.1]). There were no standardized mean differences in SVI total scores for black compared with white applicants in either cohort. There were small standardized mean differences in total SVI scores for Latino applicants and Asian applicants compared with white applicants in the 2016 cohort, but there were no differences in the 2017 cohort. There was no standardized mean difference in total SVI scores between male and female applicants in the 2016 cohort, but there was a small difference favoring female applicants in the 2017 cohort. In both cohorts, there were medium standardized mean differences in SVI total scores for FMGs compared with US-MDs and small standardized mean differences in SVI total scores for US-IMGs and DOs compared with US-MDs.

Table 3 summarizes the correlations between SVI total scores and a subset of theoretically related and unrelated selection variables. The pattern of correlations was similar across both study cohorts. As expected, the correlations between SVI total scores and USMLE Step 1 scores ($r = 0.09$ to 0.15, $P < .01$) and AOA membership ($r = 0.09$ to 0.11, $P < .01$) did not rise to the level of a practical effect or were small. There were small correlations ($r = 0.12$ to 0.21, $P < .01$) between SVI total scores and Step 2 CK scores, as well as no-to-small correlations between SVI total scores and USMLE Step 2 CS pass/fail scores ($r = -0.01$, not significant to 0.15, $P < .01$) and GHHS membership ($r = 0.12$ to 0.13, $P < .01$).

Discussion

These studies investigated initial evidence of the validity of SVI total scores and the feasibility of the SVI for use in residency selection using data from two cohorts of applicants and in research and operational settings. Other currently available selection tools (e.g., MSPE,[26] eSLOE,[13] letters of recommendation) also purport to measure interpersonal and communication skills and professionalism, but use of these tools often results in a narrow range of mostly positive or high scores/ratings that are specific to individual medical schools.

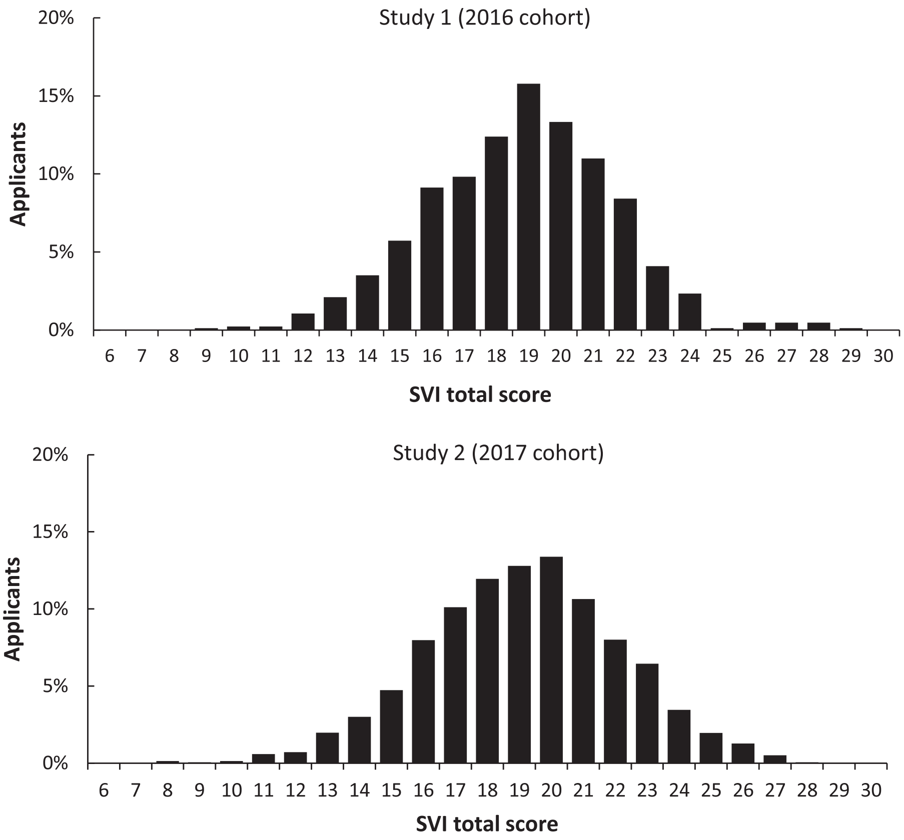

Figure 1 Comparison of the AAMC Standardized Video Interview (SVI) total score distributions in Study 1 and 2. In Study 1 (2016 cohort), 855 applicants who were considering applying to emergency medicine or who applied to pediatrics, internal medicine, or general surgery in the ERAS 2017 cycle completed the SVI in 2016. Their mean score was 18.7 with a standard deviation of 2.8. In Study 2 (2017 cohort), 3,532 applicants who were considering applying to emergency medicine for the ERAS 2018 cycle completed the SVI in 2017. Their mean score was 19.1 with a standard deviation of 3.1. Each SVI form consists of six questions, with three questions assessing knowledge of professionalism and three questions assessing interpersonal and communication skills. Trained raters used a five-point scale, from 1 (rudimentary) to 5 (exemplary), to rate the responses to each question. These question ratings were summed to create a total score ranging from 6 to 30. Abbreviations: AAMC indicates Association of American Medical Colleges; ERAS, Electronic Residency Application Service.

In contrast, in both of our studies, SVI total scores were approximately normally distributed. This wide range of scores allowed for better differentiation of candidates, which could aid program directors in their decision making. We attribute this score range to SVI raters' being objective third parties whose ratings were not influenced by personal relationships with individual applicants. Because raters were trained and used a standardized rating scale, SVI total scores had the same meaning across applicants, allowing for easy comparisons of applicants across medical schools.

Results from the two studies differ with respect to group differences in SVI total scores. In Study 1 (2016 cohort), there were no differences between black and white applicants and between women and men, but there were small differences for Latino applicants and Asian applicants compared with white applicants. In contrast, in Study 2 (2017 cohort) there was a small difference between men and women, but there were no differences by race/ethnicity. We attribute the reduction in racial/ethnic group differences to increasing the rater training time from 4 to 16 hours, with an increased emphasis on using a structured process for evaluating responses and more time dedicated to minimizing unconscious bias.

Our findings of no-to-small group differences in SVI total scores are consistent with face-to-face structured interviews in the employment domain and are likely due to the highly standardized process used to train raters and evaluate responses.[18] These findings should be interpreted in the context of standardized admissions tests in which large standardized mean differences are observed for black applicants compared with white applicants and medium standardized mean differences are observed for Latino compared with white applicants.[27] Our findings suggest that SVI total scores, when used as one piece of the residency selection process and/or in a composite, could be used to improve the diversity of the pool of applicants invited to in-person interviews.

Our results also showed small differences in SVI total scores for DOs and US-IMGs compared with US-MDs and medium differences for FMGs compared with US-MDs. These results were consistent across both studies, and more research is needed to understand them. Potential explanations could be differences in experience with video interviews, access to resources needed to prepare for the SVI, or experiences and expectations about interpersonal and communication skills and professionalism in the curriculum.

SVI total scores correlated in expected ways with other selection data. There were no-to-small correlations between SVI total scores and theoretically unrelated academic variables (i.e., Step 1 scores and AOA membership). This finding is encouraging because it suggests that SVI total scores are measuring something different from academic performance and may add valuable information to the selection process.

Findings related to theoretically related variables were mixed. There was a small correlation between SVI total scores and Step 2 CK scores. Although there is some overlap of the concepts tested, the Step 2 CK exam assesses medical knowledge and its application to patient care, which is not assessed in the SVI; therefore, it is not surprising that the correlation was small. There were also small correlations between SVI total scores and Step 2 CS scores and GHHS membership. Conceptually, we expected SVI total scores to be positively correlated with both of these variables; however, the lack of variance in Step 2 CS scores (i.e., pass or fail) may have limited our ability to detect a relationship.

Table 2
AAMC Standardized Video Interview (SVI) Total Scores in Study 1 and Study 2 by Race/Ethnicity, Gender, and Applicant Type[a]

Characteristic	No.	Mean	SD	*T*	*P* value	Standardized mean difference (*d*)[b,c,d]
Study 1 (2016 cohort, n = 855)						
Race/ethnicity[e]						
White	397	19.1	2.7	—	—	—
Black	90	19.0	2.7	0.4	.73	0.04
Latino	63	18.5	2.4	1.7	.10	0.22
Asian	218	18.1	2.9	4.4	< .001	0.38
Gender						
Male	516	18.6	2.7	—	—	—
Female	337	18.8	2.9	−0.8	.40	−0.06
Applicant type						
US-MD	441	19.3	2.5	—	—	—
DO	114	18.7	2.9	2.1	.04	0.22
US-IMG	137	18.3	2.8	3.7	< .001	0.39
FMG	161	17.4	2.9	7.4	< .001	**0.76**
Study 2 (2017 cohort, n = 3,532)						
Race/ethnicity[e]						
White	2,001	19.2	3.0	—	—	—
Black	247	19.3	3.0	−0.9	.39	−0.06
Latino	286	18.9	3.1	1.4	.16	0.09
Asian	613	19.1	3.1	0.11	.92	0.01
Gender						
Male	2,311	18.9	3.1	—	—	—
Female	1,219	19.5	3.1	−5.8	< .001	−0.21
Applicant type						
US-MD	2,062	19.6	2.9	—	—	—
DO	915	18.6	3.0	8.1	< .001	0.33
US-IMG	320	18.2	3.4	6.9	< .001	0.47
FMG	220	18.1	3.4	6.1	< .001	**0.50**

Abbreviations: AAMC indicates Association of American Medical Colleges; US-MD, attendee of an MD-granting U.S. medical school; DO, attendee of a DO-granting medical school; US-IMG, U.S. citizen attendee of an international medical school; FMG, non-U.S. citizen attendee of an international medical school.

[a] In Study 1, conducted for research purposes, applicants participated in the SVI during 2016 for the ERAS 2017 cycle. These applicants indicated interest in applying to emergency medicine or applied to the other specialties indicated in Table 1. In Study 2, the operational pilot, applicants participated in the SVI in 2017 for the ERAS 2018 cycle. The pilot was limited to applicants who indicated interest in applying to emergency medicine. Participation in the SVI was optional and free of charge in both studies.

[b] Dashes indicate the reference groups.

[c] Standardized mean difference (Cohen's *d*) = (mean of the majority group − mean of the minority group) / majority group standard deviation.

[d] Bolded values reflect medium *d*s; all other values reflect small *d*s or no difference. A *d* of 0 indicates no difference in mean score between groups. A positive *d* indicates that the majority group mean is higher than the minority group mean, and a negative *d* indicates that the minority group mean is higher than the majority group. The rule of thumb for interpreting the magnitude of the difference is that a *d* of less than 0.2 is no practical effect, 0.2 is a small effect, 0.5 is a medium effect, and 0.8 is a large effect.

[e] Individuals who self-identified as white alone were classified as white, individuals who self-identified as black alone or in combination with other races/ethnicities (including white) were classified as black, and individuals who self-identified as Latino alone or in combination with other races/ethnicities (including white) were classified as Latino. Applicants who self-identified as other racial/ethnic groups or who did not indicate race/ethnicity were not included in the race/ethnicity analyses.

The GHHS nomination process varies by school, and limited information is available about the specific criteria schools use to finalize membership. If different criteria or processes were used by schools, this could have affected our results. GHHS membership appears to measure broader constructs than interpersonal and communication skills and professionalism.

Implications

There are several potential implications of using the SVI in residency selection, such as difficulty interpreting new score types, inappropriate use of SVI total scores, and unintended consequences of altering the selection process. The EM program community and the AAMC have been taking steps to address these potential consequences.

Through training materials and outreach efforts, the AAMC has encouraged program directors to use SVI total scores cautiously while research on the meaning of these scores continues. SVI total scores are criterion-referenced: EM program directors and faculty established the scoring rubric on which the rating scale was developed by identifying several behavioral examples of each proficiency level. As such, higher scores on the SVI reflect higher levels of proficiency on the target competencies. The AAMC has also advised program directors that SVI total scores should not be used in isolation or as a cutoff. Rather, they should be interpreted in the context of other assessments that may measure similar competencies (e.g., eSLOE, MSPE) and as a complement to assessments that measure academic or technical readiness for residency (e.g., USMLE Step exam scores). Using SVI total scores cautiously and in this broader context should minimize the risk of unintended negative consequences for applicants.

Using the SVI as one part of a holistic selection process may broaden the skill set of applicants invited to in-person interviews; however, this is an empirical question that should be explored with future research. If programs use the SVI total scores to balance USMLE Step exam scores and lower their initial screening thresholds, different applicants (i.e., those who have broader skill sets) may be considered for in-person interviews. However, if programs add the SVI as another screen along with Step exam scores, it could result in programs focusing on only applicants who have both high Step exam scores and high SVI total scores. Although not our intention, one might argue that such a shift in strategies would be better than simply relying on Step exam scores alone

Table 3

Correlations Between AAMC Standardized Video Interview (SVI) Total Scores and Selection Variables in Study 1 and Study 2[a]

Variable	No.[b]	r[c]	P value
Study 1 (2016 cohort, n = 855)			
USMLE Step 1 score	782	0.15[d]	< .01
USMLE Step 2 CK score	752	0.21[d]	< .01
USMLE Step 2 CS score	560	−0.01[e]	ns
AOA membership	577	0.11[e]	< .01
GHHS membership	628	0.13[e]	< .01
Study 2 (2017 cohort, n = 3,532)			
USMLE Step 1 score	2,977	0.09[d]	< .01
USMLE Step 2 CK score	2,596	0.12[d]	< .01
USMLE Step 2 CS score	1,058	0.15[e]	< .01
AOA membership	2,700	0.09[e]	< .01
GHHS membership	2,976	0.12[e]	< .01

Abbreviations: AAMC indicates Association of American Medical Colleges; USMLE, United States Medical Licensing Examination; CK, Clinical Knowledge; CS, Clinical Skills; AOA, Alpha Omega Alpha Honor Medical Society; GHHS, Gold Humanism Honor Society; ns, not significant.

[a] For cohort characteristics and SVI total scores, see Tables 1 and 2.

[b] Number of applicants for whom data were available.

[c] The rule of thumb for interpreting the magnitude of a correlation is that $r = 0.1$ is a small effect, $r = 0.30$ is a medium effect, and $r = 0.50$ is a large effect.

[d] Values for USMLE Step 1 and Step 2 CK scores are Pearson correlations because the outcomes are continuous.

[e] Values for USMLE Step 2 CS score, AOA membership, and GHHS membership are point-biserial correlations because the outcomes are binary.

because it would likely result in more well-rounded applicants being invited to interview in person.

One unintended consequence of the SVI could be an increase in burden and cost for applicants due to the perception that time-consuming and/or expensive interview preparation is needed. Initial research shows that modest amounts of preparation time and use of the free resources provided by the AAMC and medical schools improved scores slightly compared with no preparation.[28] We recommend that the AAMC continue to provide free SVI preparation materials to advisors and applicants and to message that expensive commercial interview preparation is unnecessary for most applicants.

Finally, there are several outstanding questions about the feasibility of large-scale operational administration. The AAMC has invested considerable resources to train raters and score SVI responses. To scale the SVI to the full residency applicant pool, the AAMC may have to move to computer-based scoring. Although computer-based scoring is common in high-stakes writing assessments,[29] it is new to video interviews. More research is needed to ensure that computer-based scoring is reliable, valid, and fair. There are also unanswered questions about the cost and resources required of residency programs and medical schools to support the SVI, and about sharing applicants' video responses with program directors. Although it is not realistic to expect program directors to view all SVI videos, we felt that providing SVI total scores and the videos during the operational pilot (Study 2) was critical for program directors to understand the meaning of the scores. The AAMC allowed programs to turn off video access (as they can with applicant photographs in the ERAS application), so programs concerned about potential for implicit bias or that prefer a "closed" in-person interview process had that option. We recommend that the EM community and the AAMC revisit whether release of videos is useful going forward.

Limitations

There are several limitations to these studies. Program directors consider a wide range of data in the selection process. Our lack of access to other selection data (e.g., eSLOEs) and trainee performance outcomes prevented us from conducting a more complete evaluation of the SVI at this time. Access to other selection data would have improved our ability to establish the nomological network for the SVI[30] and evaluate the potential usefulness of the SVI. Also, we do not know whether the use of nonphysicians as raters reduced the accuracy of SVI ratings or whether the two competencies can generate reliable subscores that provide unique information.

Future research

Future research should explore the effects on SVI total scores of retaking the SVI, practice, and/or coaching. In addition, future research should examine the relationship between SVI total scores and performance outcomes during residency, as well the incremental validity of the SVI compared with other assessments designed to measure similar competencies (e.g., eSLOE, MSPE) and Step 2 CS subscores that measure related competencies. The AAMC has partnered with 17 EM programs to study the relationship between SVI total scores, locally held application data, and intern performance through at least 2020. Future research also should explore program director and applicant reactions to the SVI, including satisfaction, intended use, and perceived value. Additional outreach and communication are needed to ensure that the academic medicine community understands the SVI.

Conclusions

Findings from two studies and two ERAS cycles suggest that the SVI could enhance the current residency selection process. SVI total scores are reliable and comparable across applicants, making it possible to use these scores effectively in screening for invitations for in-person interviews. Our findings provide initial evidence of the validity of SVI total scores based on content, response processes, and relations with other variables.[21] Results also suggest that SVI total scores provide different information than what is currently available from academic metrics and that group differences are smaller than those observed on standardized tests. Use of the SVI, as part of a holistic screening process, may give program directors an opportunity to widen the skill set of applicants they invite to in-person interviews and may signal to applicants that programs value interpersonal and communication skills and professionalism.

Acknowledgments: The authors wish to acknowledge the Emergency Medicine Standardized Video Interview working group members: Ashely Alker, MD (University of California, San Diego); Andra Blomkalns, MD (Stanford University School of Medicine); Steve Bird, MD (University of Massachusetts Medical School); Mary Calderone Hass, MD (University of Michigan Health System); Nicole Deiorio, MD (Virginia Commonwealth University School of Medicine); Ramnick Dhaliwal, MD (Hennepin County Medical Center); Fiona Gallahue, MD (University of Washington); H. Gene Hern, MD (Highland Hospital); Yolanda Haywood, MD (George Washington University School of Medicine and Health Sciences); Kathy Hiller (University of Arizona College of Medicine–Tucson); Zach Jarou, MD (University of Chicago); Rahul Patwari, MD (Rush University Medical Center); Christopher Woleben, MD (Virginia Commonwealth University School of Medicine); and Richard Wolfe, MD (Harvard Medical School and Beth Israel Deaconess Medical Center).

Funding/Support: This research was funded by the Association of American Medical Colleges.

Other disclosures: D. Dunleavy is an employee and K. Dowd is a former employee of the Association of American Medical Colleges.

Ethical approval: The institutional review board of the American Institutes for Research (FWA00001666) approved Study 1 on June 2, 2016, and Study 2 on September 30, 2017. Individuals provided explicit consent for their data to be used for research purposes when they took the SVI.

Previous presentations: A subset of these data were reported at the following meetings: 2017 Council of Residency Directors in Emergency Medicine (CORD) Academic Assembly; April 29, 2017; Fort Lauderdale, Florida; 2017 Society of Academic Emergency Medicine (SAEM) Annual Meeting; May 18, 2017; Orlando, Florida; and Learn Serve Lead 2017: The AAMC Annual Meeting; November 5, 2017; Boston, Massachusetts.

Data: This manuscript was reviewed and approved for publication by the leadership of the Association of American Medical Colleges and of the National Board of Medical Examiners.

S.B. Bird is program director, Department of Emergency Medicine, and vice chair for education, University of Massachusetts Medical School, Worcester, Massachusetts.

H.G. Hern is associate clinical professor, Department of Emergency Medicine, and vice chair of education, Highland Hospital, Oakland, California.

A. Blomkalns is chair, Department of Emergency Medicine, Stanford University School of Medicine, Stanford, California.

N.M. Deiorio is associate dean for student affairs and professor, Department of Emergency Medicine, Virginia Commonwealth University School of Medicine, Richmond, Virginia.

Y. Haywood is senior associate dean for diversity and inclusion, associate dean for student affairs, and associate professor, Department of Emergency Medicine, George Washington University, Washington, D.C.

K.M. Hiller is professor and director of undergraduate education, Department of Emergency Medicine, University of Arizona College of Medicine–Tucson, Tucson, Arizona.

D. Dunleavy is director of admissions and selection research and development, Association of American Medical Colleges, Washington, D.C.

K. Dowd was a data scientist, Association of American Medical Colleges, Washington, D.C., at the time of the study.

References

1 Hamdy H, Prasad K, Anderson MB, et al. BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. Med Teach. 2006;28:103–116.

2 Emanuel EJ, Gudbranson E. Does medicine overemphasize IQ? JAMA. 2018;319:651–652.

3 Holmboe ES, Edgar L, Hamstra S. The Milestones Guidebook: Version 2016. Chicago, IL: Accreditation Council for Graduate Medical Education; 2016. http://www.acgme.org/Portals/0/MilestonesGuidebook.pdf?ver=2016-05-31-113245-103. Accessed January 14, 2019.

4 Association of American Medical Colleges. Core Entrustable Professional Activities for Entering Residency: Curriculum developers' guide. https://store.aamc.org/downloadable/download/sample/sample_id/63/. Published 2014. Accessed July 26, 2019.

5 Levinson W, Roter DL, Mullooly JP, Dull VT, Frankel RM. Physician–patient communication. The relationship with malpractice claims among primary care physicians and surgeons. JAMA. 1997;277:553–559.

6 Anhang Price R, Zyzanski SJ, Alemango AM, et al. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev. 2014;71: 522–554.

7 Berger JS, Cioletti A. Viewpoint from 2 graduate medical education deans: Application overload in the residency match process. J Grad Med Educ. 2016;8: 317–321.

8 Chen A, Shinkai K. Rethinking how we select dermatology applicants—Turning the tide. JAMA Dermatol. 2017;153:259–260.

9 Bandiera G, Abrahams C, Ruetalo M, Hanson MD, Nickell L, Spadafora S. Identifying and promoting best practices in residency application and selection in a complex academic health network. Acad Med. 2015;90:1594–1601.

10 Dunleavy D, Geiger T, Overton, R, Prescott J. Results of the 2016 Program Directors Survey: Current Practices in Residency Selection. Washington, DC: Association of American Medical Colleges; 2016. https://store.aamc.org/results-of-the-2016-program-directors-survey.html. Accessed July 26, 2019.

11 Stone C. Market Research for Standardized Video Interview (SVI) Product: Qualitative Research Findings [unpublished report]. Chicago, IL: Cheryl Stone and Associates, Ltd.; 2017.

12 Prober CG, Kolars JC, First LR, Melnick DE. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. Acad Med. 2016;91:12–15.

13 Love JN, Smith J, Weizberg M, et al; SLOR Task Force. Council of Emergency Medicine Residency Directors' standardized letter of recommendation: The program director's perspective. Acad Emerg Med. 2014;21: 680–687.

14 Keim SM, Rein JA, Chisholm C, et al. A standardized letter of recommendation for residency application. Acad Emerg Med. 1999;6:1141–1146.

15 Shah SK, Arora S, Skipper B, Kalishman S, Timm TC, Smith AY. Randomized evaluation of a web based interview process for urology resident selection. J Urol. 2012;187:1380–1384.

16 Salgado JF, Viswesvaran C, Ones DS. Predictors used for personnel selection: An overview of constructs, methods, and techniques. In: Anderson NA, Ones DS, Sinangil HK, Viswesvaran C, eds. Handbook of Industrial and Organizational Psychology. Thousand Oaks, CA: SAGE Publications; 2001;165–199.

17 Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education? A systematic review. Med Educ. 2016;50:36–60.

18 Lavashina J, Hartwell CJ, Morgenson FP, Campion MA. The structured employment interview: Narrative and quantitative review of the research literature. Pers Psychol. 2014;67:241–293.

19 Van Iddekinge CH, Raymark PH, Roth PL, Payne HS. Comparing the psychometric characteristics of ratings of face-to-face and videotaped structured interviews. Int J Sel Assess. 2006;14:347–359.

20 Blacksmith N, Willford JC, Behrend TS. Technology in the employment interview: A meta-analysis and future research agenda. Pers Assess Decis. 2016;2(1):article 2.

21 American Educational Research Association; American Psychological Association; National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.

22 Accreditation Council for Graduate Medical Education. Milestones by specialty. https://www.acgme.org/What-We-Do/Accreditation/Milestones/Milestones-by-Specialty. Accessed February 28, 2019.

23 Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979;86:420–428.

24 McCormack WT, Lazarus C, Stern D, Small PA Jr. Peer nomination: A tool for identifying medical student exemplars in clinical competence and caring, evaluated at three

medical schools. Acad Med. 2007;82:1033–1039.

25 Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

26 Hom J, Richman I, Hall P, et al. The state of medical student performance evaluations: Improved transparency or continued obfuscation? Acad Med. 2016;91:1534–1539.

27 Sackett PR, Shen W. Subgroup differences on cognitively loaded tests in contexts other than personnel selection. In: Outtz JL, ed. Adverse Impact: Implications for Organizational and Staffing and High Stakes Selection. New York, NY: Taylor and Francis Group; 2010:323–346.

28 Factors affecting Standardized Video Interview performance: Preparation elements and the testing environment. EM Resident. April 17, 2018. https://www.emra.org/emresident/article/svi-study-results. Accessed January 14, 2019.

29 Dikli S. An overview of automated scoring of essays. J Tech Learn Assess. 2006;5:1–35.

30 Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52:281–302.

Teaching and Learning Moments
Reflections on a New Curriculum

Dust settled through the late September afternoon, with blood orange rays of sun sliding diagonally through the living room window. The dust had been kicked up by children racing in between moving boxes stacked like skyscrapers in their own pretend city. The new house meant new hiding places, new neighborhood kids to play with, and new windows for the family dog to bark from protectively. I shook J.M.'s hand and took a seat at the dining room table. As he apologized for the clutter around us, I felt an uplifting sense of youthful innocence in the air. The feeling was blindsided and sunk by one realization: His kids knew they had moved because their dad needed to be closer to the hospital.

He told me everything. More than a decade of every type of imaginable pain from a myriad of chronic comorbidities was shared with a complete stranger. I could feel his voice shaking when detailing particularly difficult times, and the same voice deepening with strength as he worked to suppress the tears brewing just underneath his eyelids. From emergency room visits to expert referrals, from medication trials to being told it was all in his head, a common theme emerged over the years: a surplus of symptoms and a scarcity of answers. He opened his personal life and shared the medical challenges in a way I had never believed possible. I stared down at the glass of ice water in my hands, desperately hoping the right thing to say would somehow appear in the tiny bubbles frozen in the ice cubes. What could I do? What could I say? This was only my first week of medical school, and the expected textbooks and lecture slides were nowhere to be found. So, I did the only thing I knew how to do. I listened.

Fourteen months into medical school, on the first day of the family medicine clerkship, I saw J.M.'s name on my patient list for the afternoon. Flashbacks of the tearful conversation about debilitating pain and frustration from unanswered questions came to me as I was about to knock on the door to the room. I paused, knocked twice, and entered the clinic room, where I was met by a tired but welcoming smile from J.M. Although I had not had much to offer at our initial home encounter, absorbing details of his medical history, symptoms that caused him trouble, and his personal aspirations and career plans helped to build rapport and cultivate trust in the clinical setting. He knew how early I was in my training. He knew I would not have answers to all of his questions, but he still had faith that I was on his team. Sharing the intimate and emotional details of his life in the past and feeling truly heard made for the type of human connection you cannot find in textbooks.

So, we talked. The standard set of patient questions quickly flowed into a more natural conversation. He was eager to share his progress, and his children chimed in with their own updates throughout the visit. We shared so much in a short period of time, and before I knew it, I was back in the team room writing up my clinical note. My erratic typing was interrupted by the sudden realization that in just over a year, I had gone from asking about this man's health to playing a direct role in his care. Our former visit allowed me to look at the current one through a more compassionate and holistic lens. I have heard that doctors have one moment in their training that sticks out, a moment when they feel like a real medical provider for the first time, sometimes through an amalgamation of different experiences. Such raw, career-shaping experiences are certainly multidimensional, without a score or percentile to their name. Early clinical immersion pushes us to grow as compassionate caregivers and provides us with the experiences to cling to when challenges arise in the future—invaluable lessons in what it means to be truly present with a patient and how that fosters rapport building and patients' buy-in regarding their own health. Most of all, this experience introduced me to the beauty of having a patient feel genuinely cared for, which will undoubtedly guide my practice of medicine in years to come.

I had my first "This must be what it's all about" clinical moment just over a year into medical school. The emotions that were precipitated by meeting and then caring for J.M. will remain with me moving forward because for the very first time, not only did I feel like a doctor, I felt like somebody's doctor.

Author's Note: The name and identifying information in this essay have been changed to protect the identity of the individuals described.

Nicholas W. Eyrich, MS

N.W. Eyrich is a third-year medical student, University of Michigan Medical School, Ann Arbor, Michigan; email: eyrichn@med.umich.edu.